

Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination

Xueping Zhou,^{1,4} Yule Liu,^{1,5} Lee Calvert,² Claritza Munoz,² G. William Otim-Nape,^{1,6} David J. Robinson¹ and Bryan D. Harrison^{1,3}

¹ Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

² Centro Internacional de Agricultura Tropical, AA6713, Cali, Colombia

³ Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, UK

⁴ Institute of Biotechnology, Zhejiang Agricultural University, Hangzhou 310029, China

⁵ Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China

⁶ Namulonge Agricultural and Animal Production Research Institute, PO Box 7084, Kampala, Uganda

Geminivirus isolates associated with the epidemic of severe cassava mosaic disease in Uganda were studied and compared with virus isolates from the part of Uganda outside the epidemic area, and with African cassava mosaic virus (ACMV) and East African cassava mosaic virus (EACMV). Isolates of a novel type [the Uganda variant (UgV)] were detected in severely affected plants from the epidemic area, whereas those from plants outside the epidemic area were typical of ACMV. The complete nucleotide sequences of DNA-A of UgV (2799 nt) and of a Tanzanian isolate of EACMV (2801 nt) were determined and are extremely similar, except for the coat protein (CP) gene. The CP gene of UgV has three distinct regions: the 5' 219 nt are 99% identical to EACMV (only 79% to ACMV); the following 459 nt are 99% identical to ACMV (75%

to EACMV); and the 3' 93 nt are 98% identical to EACMV (76% to ACMV). UgV DNA-A therefore is considered to have arisen by interspecific recombination of EACMV and ACMV. Despite the hybrid nature of their CP, UgV isolates were indistinguishable from ACMV in tests with 20 monoclonal antibodies (MAbs), including seven which reacted with ACMV but not EACMV. The discontinuous epitopes detected by these seven MAbs must involve amino acids which lie in the central part of the CP (residues 74–226) and which differ in ACMV and EACMV. UgV isolates were detected in severely mosaic-affected plants from all 11 widely separated locations sampled. The probable role of recombination in geminivirus evolution in the short to medium term is discussed.

Introduction

Mosaic disease of cassava, responsible for enormous (> £1 billion) annual crop losses in Africa (Fargette *et al.*, 1988) and the Indian subcontinent, is caused by whitefly-transmitted geminiviruses. Virus isolates fall into three categories on the basis of their reactions with monoclonal antibodies (MAbs) raised against purified virus particles (Thomas *et al.*, 1986;

Harrison & Robinson, 1988; Swanson & Harrison, 1994). Comparisons of the nucleotide sequences of the complete genomes (DNA-A and DNA-B) of isolates representing two of these categories (Stanley & Gay, 1983; Hong *et al.*, 1993), and of part of DNA-A of an isolate in the third category, indicated that the isolates are not closely related and should be regarded as three separate species, which were named African cassava mosaic virus (ACMV), East African cassava mosaic virus (EACMV) and Indian cassava mosaic virus (ICMV) (Hong *et al.*, 1993). Results of extensive MAb-based serological tests on mosaic-affected cassava from 20 countries have shown that the three viruses have different, scarcely overlapping, distributions: ACMV in Africa west of the Rift Valley and South Africa; EACMV in Africa east of the Rift Valley and

Author for correspondence: Bryan D. Harrison (at the Scottish Crop Research Institute). Fax +44 1382 562 426.
e-mail djrobi@scri.sari.ac.uk

The EMBL accession numbers of the sequences reported in this paper are Z83252–Z83258.

Madagascar; and ICMV in India and Sri Lanka (Swanson & Harrison, 1994).

Cassava mosaic disease was reported in Uganda more than half a century ago (Martin, 1928), and samples collected in the 1980s from the Kampala area contained a virus serologically identical to ACMV (Swanson & Harrison, 1994 and unpublished results). However, in about 1988, an extremely severe form of mosaic disease appeared in the north-east of the country and has spread southwards and westwards, advancing about 20 km per year. It has devastated crops, caused cassava cultivation to be abandoned and resulted in starvation in areas where cassava was formerly the staple food (Otim-Nape *et al.*, 1996). In this paper, we describe nucleotide sequence and serological evidence that a geminivirus, which was found consistently in severely affected plants from the area of the epidemic, is a novel form that has arisen by recombination between EACMV and ACMV.

Methods

■ **Virus sources.** Cassava (*Manihot esculenta*) hardwood cuttings were collected from symptom-expressing plants in fields in Uganda inside and outside the area affected by the severe mosaic epidemic. The cuttings were taken to Dundee, planted in previously sterilized potting compost and the resulting plants were kept at 20–30 °C in a containment glasshouse under licence from the Scottish Office Agriculture, Environment and Fisheries Department. Other plants were grown in the containment glasshouse from cuttings obtained from mosaic-affected cassava plants growing near Msabaha in coastal Kenya and near Dar es Salaam, Tanzania, which are areas where EACMV occurs (Swanson & Harrison, 1994).

The following virus isolates (with the symptoms shown by their cassava source plants) were studied in most detail:

(a) **From the area of the severe mosaic epidemic**

- C39 (very severe mosaic) – Otuboi, Soroti District
- C70 (severe mosaic) – Mbarara District
- C74 (very severe mosaic) – Barrack, Mbarara District
- C100 (very severe mosaic) – Namulonge, Kampala District.

(b) **From outside the epidemic area (all Kampala District)**

- MBE6 (mild mosaic) (= ACMV-U2)
- SBE1 (moderate mosaic)
- SBE2 (moderate mosaic)
- SBE5 (mild mosaic)
- SBE6 (moderate mosaic) (= ACMV-U1)
- SBE11 (severe mosaic).

(c) **From outside Uganda**

- EACMV-K2B – Msabaha, coastal Kenya (= EACMV-K)
- EACMV-Dar6 – Dar es Salaam, Tanzania (= EACMV-T)
- EACMV-Dar7 – Dar es Salaam, Tanzania (= EACMV-T2)
- EACMV-YG – Rumphi, Malawi (Hong *et al.*, 1993)
- ACMV-K – western Kenya (Stanley & Gay, 1983).

■ **Epitope profiles.** The reaction of extracts of cassava leaves with a panel of MAbs was determined by triple antibody sandwich ELISA, essentially as described by Muniyappa *et al.* (1991). Leaves were extracted in 10–20 vol. of the Tris–sulphite-containing buffer 2 of

Macintosh *et al.* (1992). Murine MAbs were raised against purified particles of ACMV (Thomas *et al.*, 1986), ICMV (Aiton & Harrison, 1989) or okra leaf curl virus (OLCV) from Ivory Coast (Swanson & Harrison, 1993). The MAbs were used as tissue culture supernatant fluids diluted 1:4. Absorbance at 405 nm was recorded after 1–2 h incubation with substrate at room temperature, followed by overnight incubation at 4 °C.

■ **PCR.** Nucleic acid samples were obtained from plant leaves by method A of Harrison *et al.* (1997) or by the method of Gilbertson *et al.* (1991). Three pairs of primers were used to amplify the whole DNA-A of isolates C39 and C74, and of a Tanzanian isolate of EACMV (EACMV-T), as three overlapping fragments: UV-AL1/F1 and UV-AL1/R1; UV-AL1/F2 and UV-CP/R; and UV-AL3/F and UV-AL1/R2. DNA-A of isolate C100 was amplified as two fragments with the degenerate primer pairs PAL1v1978/PAR1c715 and PAR1v722/PAL1c1960 [after Rojas *et al.* (1993)]. A fragment including the intergenic region and the AV2 and coat protein (CP) genes of isolates C70, C74 and EACMV-K was amplified with primers UV-AL1/F2 and UV-CP/R. The intergenic region and the AV2 gene of ACMV-U1 were amplified with primers ACMV-AL1/F and ACMV-AR0/R and the CP gene of ACMV-U2 was amplified with primers ACMV-CP/F2 and ACMV-CP/R2. The sequences of the primers are shown in Table 1. Reaction mixtures were as in Harrison *et al.* (1997). The first reaction cycle comprised 1 min at 94 °C, 2 min at 52 °C and 3 min at 72 °C, and was followed by 35 cycles of 1 min at 94 °C, 1 min at 52 °C and 80 s at 72 °C, and a final 5 min at 72 °C.

■ **Sequence determination and analysis.** Except for those from isolate C100, PCR products were recovered, after electrophoresis in a 1% agarose gel, with the Wizard PCR Preps DNA purification system (Promega) and were cloned in the pT7Blue T-vector (Novagen). Plasmids were purified using a QIAprep Spin plasmid kit (Qiagen) and were sequenced on an ABI model 373A Stretch DNA sequencer (Perkin Elmer) using the PRISM Ready Reaction DyeDeoxy Terminator Cycle sequencing kit. Universal M13 (–40) and T7 promoter primers were used, as well as one additional internal primer, shown in Table 1. All residues in the sequences were confirmed by their occurrence in different isolates and/or by sequencing in both directions and/or by comparison of duplicate clones. Wherever clones overlapped, their sequences were identical. PCR products from isolate C100 were cloned in pCR-Script Amp SK(+) (Stratagene). Sequence data were obtained from these clones and from subclones in pBluescript II KS(+) (Stratagene) using a Sequenase version 2.0 DNA sequencing kit (United States Biochemical).

Sequence data were assembled and analysed with the aid of the Wisconsin Package version 8.1 programs (Anon., 1994). PILEUP was used to generate multiple alignments, which were further optimized manually.

Results

Observations on diseased cassava plants

Cuttings collected in 1995 from mosaic-affected cassava plants (mostly cv. Ebwanateraka) growing in southern Uganda in areas not reached by the severe mosaic epidemic were planted at Dundee alongside cuttings obtained from very severely affected plants growing within the area of the epidemic. Many leaves on shoots produced by the pre-epidemic cuttings developed mild to moderately severe mosaic, but other leaves on the same shoots were symptomless (Fig. 1). In contrast, most shoots produced by cuttings from the epidemic area were weak and stunted, and almost all leaves on them were small and distorted with severe mosaic (Fig. 1).

Table 1. Primers used for PCR

Primer designation*	Primer sequence (5' to 3')†	Position in DNA‡
Primers used for cloning:		
UV-AL1/F1	TGTCTTCTGGGACTTGTGTG	2057–2076 ^a
UV-AL1/R1	AACCTATCCCCGATGCTCAT	2595–2576 ^a
UV-AL1/F2	GTAATTGGGAAAAGGCCTCT	2546–2565 ^a
UV-CP/R	GTTACGGAGCAACATGCAAT	1140–1121 ^a
UV-AL3/F	TACACATGCCTCRAATCCTG	1033–1052 ^a
UV-AL1/R2	CTCCGCCACAACTTACGTT	2119–2100 ^a
ACMV-AL1/F	GCGGAATCCCTAACATTATC	1985–2004 ^b
ACMV-AR0/R	GCTCGTATGTATCCTCTAAGGCCTG	233–209 ^b
ACMV-CP/F2	ACATACGAGCCCAGTACGTT	224–243 ^b
ACMV-CP/R2	CCAACACTATTGAGTGCAAG	1119–1100 ^b
PAL1v1978§	GCATCTGCAGGCCACATYGTCTTYCCNGT	
PAR1c715§	GATTTCTGCAGTTDATRITYTCRTCCATCCA	
PAR1v722§	ATATCTGCAGGGNAARATHHTGGATGGA	
PAL1c1960§	TGGACTGCAGACNNGNAARACNATGTGGGC	
Additional primer used for sequencing:		
UV-CP/F2	GCGTCAACATGTGGGATCCATT	164–185 ^a
Primer used for specific detection of UgV:		
ACMV-CP/R3	TGCCTCCTGATGATTATATGTC	947–926 ^b
Primers for ACMV DNA-B:		
ACMV-BR1/F	TGTATAGTATCAGGAAGCAG	436–455 ^c
ACMV-BR1/R	GGTGCTCAATCGCTTATAA	1036–1017 ^c
ACMV-BL1/F	CTATAATGCTTTCAATGTCC	1273–1292 ^c
ACMV-BL1/R	CAATTGATGATAACGTACCA	1788–1769 ^c

* Designations containing /F or v denote forward primers and those containing /R or c are reverse primers.

† R represents A or G; Y represents C or T; D represents A, G or T; H represents A, C or T; N represents A, C, G or T.

‡ Residue numbers refer to the sequences of: *a*, DNA-A of isolate C39 (this paper); and *b* and *c*, DNA-A and DNA-B, respectively, of ACMV (Stanley & Gay, 1983) renumbered to make residue 1 the A in the conserved nonanucleotide TAATATTAC in the intergenic region (originally residue 153).

§ Primers are named according to Rojas *et al.* (1993).

Thus, the two main kinds of symptom observed in field conditions in Uganda were mimicked in vegetative progeny plants in the glasshouse at Dundee.

Four plants from outside the epidemic area of Uganda yielded DNA fragments with sequences typical of ACMV when primers ACMV-AL1/F and ACMV-AR0/R were used to amplify DNA from leaf extracts by PCR. In contrast, four plants from the epidemic area of Uganda, and single plants from coastal Kenya and from coastal Tanzania, each yielded DNA fragments with sequences atypical of ACMV when primers UV-AL1/F1 and UV-AL1/R1 were used. These primers did not amplify any DNA fragment from extracts of leaves from mosaic-affected pre-epidemic plants. Determination and analysis of the nucleotide sequences in these and other cloned DNA fragments are described below.

Comparison of complete DNA-A sequences

The complete nucleotide sequence of DNA-A was determined for isolates C39 and C74, and an almost complete

sequence was determined for isolate C100, all of which were obtained from very severely affected Ugandan cassava plants. Also, because only a partial DNA-A sequence was available for EACMV (Hong *et al.*, 1993), the complete sequence was determined for EACMV-T. These sequences were compared with one another and with the sequence of ACMV-K DNA-A published by Stanley & Gay (1983).

The C74 sequence and the provisional C100 sequence were > 98% identical to that of C39, although their source plants were collected up to 300 km apart (Mbarara, Kampala and Soroti Districts). The total sequence lengths were 2799 nt for isolates C39 (EMBL Z83257) and C74, and 2801 nt for EACMV-T (EMBL Z83256), as compared with 2779 nt for ACMV-K. Comparison of these three kinds of complete sequence revealed a greater percentage identity for C39 and EACMV-T (92%), than for either C39 and ACMV-K (72%) or EACMV-T and ACMV-K (68%).

Table 2 lists the positions of the open reading frames (ORFs) in DNA-A and the sizes of the putative translation



Fig. 1. (a) Plants grown from cassava cuttings collected from (left) two severely diseased plants from the area of the epidemic, and (right) two plants from outside the epidemic area. Note extreme stunting and severe mosaic shown by plants in the left-hand pot and the much lesser degree of stunting of plants in the right-hand pot. (b) Detail of symptoms in leaves from plants shown in (a). Left, leaf from extremely stunted plant from the epidemic area showing severe mosaic. Centre and right, leaves from plants from outside the epidemic area, showing moderate and mild mosaic, respectively.

Table 2. Open reading frames in DNA-A of EACMV, ACMV and isolate C39

ORF	EACMV-T		C39		ACMV-K	
	Position (nt)	No. of encoded amino acids	Position (nt)	No. of encoded amino acids	Position (nt)	No. of encoded amino acids
AV2	174–527	118	172–525	118	134–472	113
AV1 (CP)	334–1104	257	332–1102	257	294–1067	258
AC1 (Rep)	2644–1583	354	2642–1566	359	2604–1531	358
AC2	1656–1252	135	1654–1250	135	1619–1215	135
AC3	1508–1107	134	1506–1105	134	1471–1070	134
AC4	2487–2257	77	2485–2255	77	2450–2151	100

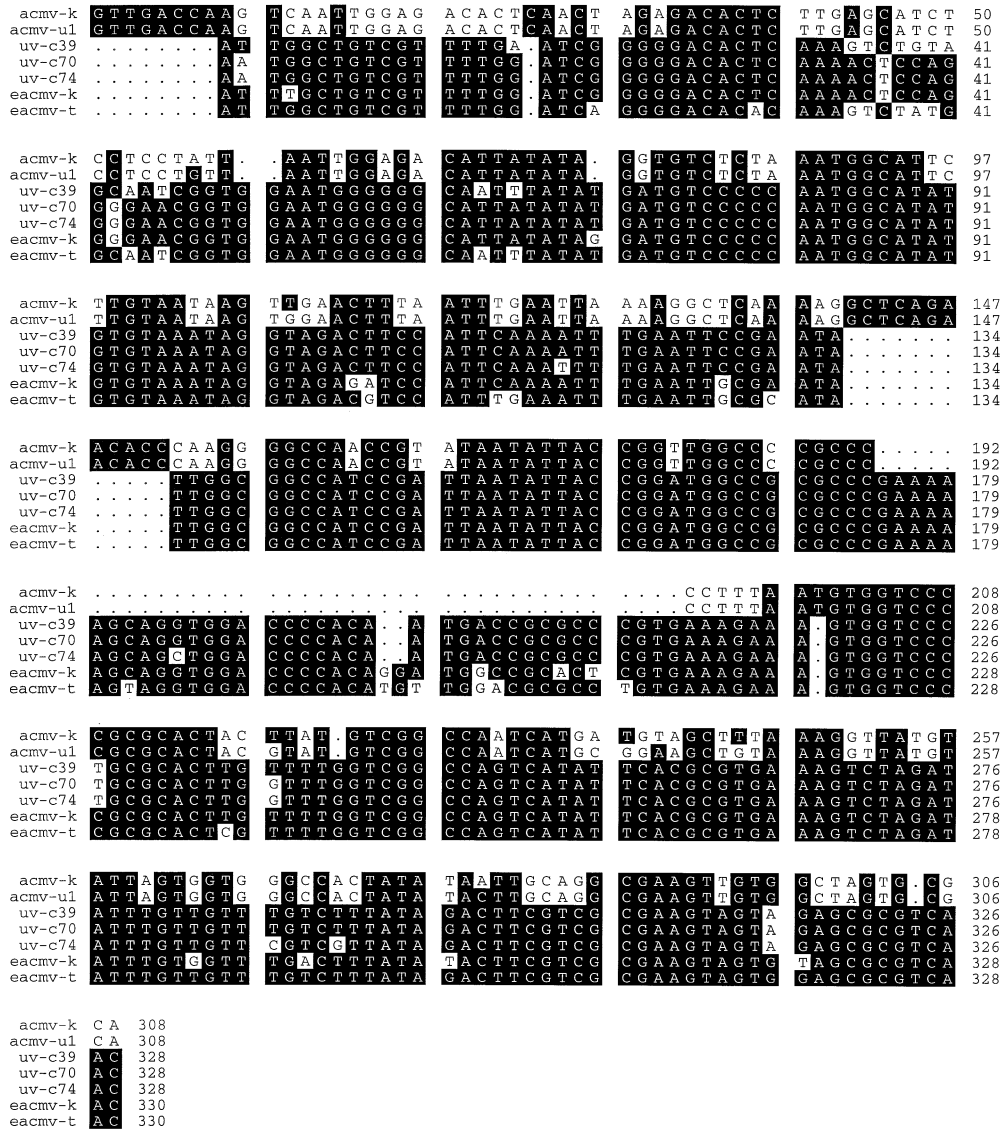


Fig. 2. Comparison of nucleotide sequences of the intergenic region of virus isolates from inside (C39, C70, C74) and outside (ACMV-U1) the epidemic area in Uganda, together with those of ACMV-K, EACMV-K and EACMV-T.

Table 3. Percentage amino acid sequence identities of putative products of ORFs in DNA-A of EACMV, ACMV and isolate C39

ORF	Viruses compared		
	C39 and EACMV-T	C39 and ACMV-K	EACMV-T and ACMV-K
AV2	95	62	60
AV1(CP)	89	91	81
AC1(Rep)	97	69	69
AC2	90	63	64
AC3	90	69	70
AC4	91	31	30

products for the three isolates. The nucleotides are numbered in the same way in each sequence, with nt 1 being A in the conserved nonanucleotide TAATATTAC found in the large intergenic region of all geminiviruses. The arrangement of ORFs in DNA-A of C39 and EACMV-T was similar to that in DNA-A of other whitefly-transmitted geminiviruses from the Old World. The positions and sizes of the ORFs were almost the same for the two viruses, the only noteworthy difference being 15 fewer nucleotides at the 3' end of the AC1 gene of EACMV-T. Sequencing of a second independent clone of EACMV-T gave the same result. However, isolates of EACMV were found to differ in the size of their AC1 gene; in EACMV-K it was the same size as in C39. When ACMV-K was compared with C39 or EACMV-T, all the ORFs differed somewhat in position and four of the six gene products differed in size (Table 2).

Comparison of intergenic regions

The intergenic region is the part of DNA-A that differs to the greatest extent among whitefly-transmitted geminiviruses. Great similarity between the intergenic regions of two geminivirus isolates constitutes strong evidence that they are closely related (Rybicki, 1994; Padidam *et al.*, 1995). Sequences of the intergenic regions of three virus isolates from cassava from the epidemic area in Uganda, of one isolate from outside the epidemic area (ACMV-U1), of EACMV-T and EACMV-K, and the equivalent published sequence of ACMV-K, are compared in Fig. 2. The intergenic regions of the three isolates (C39, C70 and C74) from the epidemic area are identical in length (326 nt) and very similar in sequence. Those of the two EACMV isolates have the same length (328 nt) and strong

sequence similarity to one another. The consensus sequence for the C isolates is 95% identical to that of EACMV-K. The ACMV isolates, in contrast, although very similar to one another, differed considerably from the others in length (308 nt; 64–66 insertions/deletions) and sequence. These comparisons illustrate the great similarity of the intergenic region of C isolates to that of EACMV isolates, and the difference from ACMV isolates. It was also found that the intergenic region of the C isolates is more like that of the EACMV isolates included in Fig. 2 than that of the EACMV isolate from Malawi studied by Hong *et al.* (1993), which is more divergent in the 5' 122 nt but similar in the remainder of the sequence (not shown).

As with the intergenic regions of other whitefly-transmitted geminiviruses, those of the C isolates and EACMV contain various special elements. These include the conserved non-nucleotide motif (TAATATTAC), bounded by sequences capable of forming a stem-loop structure (nt 138–170 in Fig. 2), and a TATA motif at nt 67–70. In addition, the sequence ATGGGCTG at nt 1–8 in isolate C39 is followed by AATCGGTG at nt 44–51, AATGGGGG at nt 53–60 and the nearly complementary sequence TAACCCCC at nt 84–77. In other geminiviruses, nearly reiterated sequences of this kind interact with the AC1 gene product during viral DNA replication (Fontes *et al.*, 1992; Argüello-Astorga *et al.*, 1994).

Affinities of virus-encoded proteins

The close similarity between C39 and EACMV-T is very evident when the amino acid sequences of their putative gene products are compared with one another and with those of ACMV-K (Table 3). Of the six gene products, only CP has less

acmv-k	G T G A A A A G G T	T T T A C A G G T T	G A A T C A T C A C	G T G A C A T A C A	A T C A T C A G G A	650
acmv-u2	G T G A A A A G G T	T T T A C A G G T T	G A A T C A T C A C	G T G A C A T A T A	A T C A T C A G G A	650
uv-c39	G T G A A A A G G T	T T T A C A A G C T	G A A T C A T C A C	G T G A C A T A T A	A T C A T C A G G A	647
uv-c70	G T G A A A A G G T	T T T A C A A G C T	G A A T C A T C A C	G T G A C A T A T A	A T C A T C A G G A	647
uv-c74	G T G A A A A G G T	T T T A C A A G C T	G A A T C A T C A C	G T G A C A T A T A	A T C A T C A G G A	647
eacmv-k	G T T A A G A G G T	T T T T T A G G A T	C A A T A A T C A T	G T T G T G T A T A	A T C A T C A G G A	647
eacmv-t	G T C A A G A G G T	T T T T T A G G A T	T A A T A A T C A T	G T A G T G T A T A	A T C A T C A G G A	647
acmv-k	G G C A G G G A A G	T A T G A G A A T C	A C A C A G A G A A	T G C T T T G C T T	C T G T A C A T G G	700
acmv-u2	G G C A G G G A A G	T A T G A G A A T C	A C A C A G A G A A	T G C T T T G C T T	C T G T A C A T G G	700
uv-c39	G G C A G G G A A G	T A T G A G A A T C	A C A C A G A G A A	T G C G T T G T T A	T T G T A T A T G G	697
uv-c70	G G C A G G G A A G	T A T G A G A A T C	A C A C A G A G A A	T G C G T T G T T A	T T G T A T A T G G	697
uv-c74	G G C A G G G A A G	T A T G A G A A T C	A C A C A G A G A A	T G C G T T G T T A	T T G T A T A T G G	697
eacmv-k	A C A G G C C A A G	T A T G A G A A T C	A T A C T G A G A A	T G C G T T G T T A	T T G T A T A T G G	697
eacmv-t	A C A G G C C A A G	T A T G A G A A T C	A T A C T G A G A A	T G C G T T G T T A	T T G T A T A T G G	697
acmv-k	C A T G T A C T C A	T G C C T C C A A T	C C T G T A T A T G	C G A C G T T G A A	A A T A C G T A T A	750
acmv-u2	C A T G T A C G C A	T G C C T C C A A T	C C T G T A T A T G	C G A C G T T G A A	A A T A C G T A T A	750
uv-c39	C A T G T A C A C A	T G C C T C G A A T	C C T G T G T A T G	C T A C G C T G A A	A A T A C G C A T C	747
uv-c70	C A T G T A C A C A	T G C C T C G A A T	C C T G T G T A T G	C G A C G C T G A A	A A T A C G C A T C	747
uv-c74	C A T G T A C A C A	T G C C T C G A A T	C C T G T G T A T G	C T A C G C T G A A	A A T A C G T A T C	747
eacmv-k	C A T G T A C A C A	T G C C T C G A A T	C C T G T G T A C G	C T A C G C T G A A	A A T A C G C A T T	747
eacmv-t	C A T G T A C A C A	T G C C T C G A A T	C C T G T G T A C G	C T A C G C T G A A	A A T A C G C A T C	747
acmv-k	T A C T T C T A T G	A C A G T A T T G G	C A A T T A A	777		
acmv-u2	T A C T T C T A T G	A C A G T A T T G C	C A A T T A A	777		
uv-c39	T A T T T T T A T G	A T G C A G T G A C	A A A T T A A	774		
uv-c70	T A T T T T T A T G	A T G C A G T G A C	A A A T T A A	774		
uv-c74	T A T T T T T A T G	A T G C A G T G A C	A A A T T A A	774		
eacmv-k	T A T T T C T A T G	A T G C A G T G A C	A A A T T A A	774		
eacmv-t	T A T T T C T A T G	A T G C A G T G A C	A A A T T A A	774		

Fig. 3. Comparison of nucleotide sequences of the CP gene of virus isolates from inside (C39, C70, C74) and outside (ACMV-U2) the epidemic area in Uganda, together with those of ACMV-K, EACMV-K and EACMV-T.

Table 4. Percentage nucleotide and encoded amino acid sequence identities of three regions of the coat protein gene of isolate C39 with isolates EACMV-T and ACMV-U2

Numbers are percentage nucleotide sequence identity. Numbers in parentheses are percentage amino acid sequence identity.

Isolate compared	Region of C39 gene (nt no.)		
	1–219	220–678	679–771
EACMV-T	99 (100)	75 (82)	98 (100)
ACMV-U2	79 (82)	99 (99)	76 (90)

than 90% sequence identity when C39 and EACMV-T are compared, whereas CP is the only product to have > 70% sequence identity when ACMV-K is compared with either of the other two viruses. These data confirm that EACMV and ACMV are at least as distantly related to one another as are other distinct species of whitefly-transmitted geminiviruses, a conclusion reached previously by Hong *et al.* (1993) on the basis of comparisons including only about half of the DNA-A sequence of a Malawian isolate of EACMV.

Novel kind of geminiviral coat protein

In Table 3, the data for comparison of the CP of C39 and other viruses seem anomalous. In general, the CP of whitefly-transmitted geminiviruses is conserved to a greater extent than any other gene product, as for example between EACMV-T and ACMV-K. However, the CP of C39 is slightly less similar to that of EACMV-T than are the other gene products and, in addition, is more similar to the CP of ACMV-K than to that of EACMV-T. To ascertain whether or not this result is consistent for virus isolates from the epidemic area, nucleotide sequences of CP genes were determined for two other such isolates, for an isolate from the pre-epidemic zone (ACMV-U2) and for EACMV-K; these sequences are compared in Fig. 3. The sequences of isolates C39, C70 and C74 differ in only six out of 771 nucleotides. Plainly, the C39 CP gene is fully representative of isolates from the epidemic area. Similarly, the CP genes of ACMV-K and ACMV-U2, isolates collected in different countries and with an interval of at least 15 years between the two collections, are very like one another, as are the CP genes of EACMV-T and EACMV-K. The genes are therefore of three kinds: the C isolates, EACMV and ACMV.

However, closer inspection of Fig. 3 shows that the sequence of the CP gene of the C isolates can be divided into three parts. The 5' part is extremely similar to that of EACMV but different from ACMV, whereas a large central part is very like ACMV but less like EACMV, and towards the 3' terminus the CP gene is again similar to EACMV and different from ACMV. These features are quantified in Table 4, where data

for both nucleotide sequence identity and amino acid sequence identity are given. These data are strong evidence that the CP gene of the C isolates has arisen by recombination between EACMV and ACMV. Because the C isolates differ consistently, and in a characteristic way, from both EACMV and ACMV, they will be referred to subsequently as the Uganda variant (UgV).

The precise sites of recombination cannot be identified from the sequence evidence. However, the 5' site must be between nt 198 and 221 of the C39 CP gene, a 24 nt region where the sequences of ACMV and EACMV are identical (Fig. 3). The 3' site is at nt 673–680 of the C39 gene, where there is an identical 8 nt sequence in ACMV and EACMV. The sequence around the 3' site is different from that around the 5' site.

Epitope profiles

Twenty MAbs were tested for their reactivity with UgV isolates, EACMV and ACMV. Six of these MAbs (SCR 21, 22, 25, 27, 29 and 33) can detect ACMV but not EACMV or ICMV, three (SCR 54, 56 and 66) detect only ICMV, one (SCR 53) detects ICMV and ACMV, two (SCR 18 and 20) detect all three viruses and the remaining eight detect most or all isolates of ACMV and EACMV (Swanson & Harrison, 1993, 1994). In these tests, the Ugandan isolates from outside the epidemic area and the UgV isolates had epitope profiles typical of ACMV, whereas EACMV-K and EACMV-T2 gave reactions typical of EACMV (Fig. 4). The few differences between ACMV and individual UgV isolates in the strength of reaction with a particular MAb may reflect uncontrolled differences in virus antigen concentration in the leaf extracts or minor genetic variation.

The reactivity of the UgV isolates with MAbs SCR 21, 22, 25, 27, 29, 33 and 53 indicates that the epitopes so detected, which are all discontinuous (Swanson, 1992), must involve amino acids in the middle (residues 74 to 226) of the viral CP, where UgV CP and ACMV CP are identical but differ from EACMV CP. Thus although only 16% of their DNA-A is ACMV-like and 84% EACMV-like, the UgV isolates would be grouped with ACMV in routine antigenic analyses of the type described by Swanson & Harrison (1994).

Virus discrimination using PCR

In the initial phases of this work, UgV isolates (including C39, C70 and C74) were distinguished from ACMV because they were detected during PCR with primers UV-ALI/F1 and UV-ALI/R1 but not with the ACMV-specific primers, ACMV-ALI/F and ACMV-AR0/R. However, primers UV-ALI/F1 and UV-ALI/R1 also detected isolates of EACMV. To distinguish the UgV isolates from EACMV, an additional primer, ACMV-CP/R3, was designed. In combination with primer UV-ALI/F1, this primer detected UgV isolates but not EACMV. DNA was amplified both by primers UV-ALI/F1

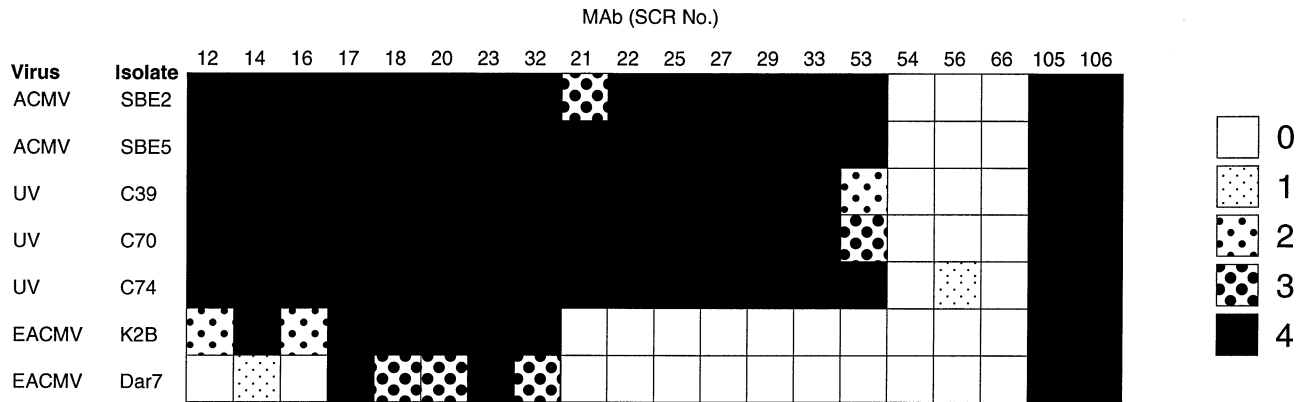


Fig. 4. Comparison of reactions with 20 geminivirus MAbs of virus isolates from inside (C39, C70, C74) and outside (SBE2, SBE5) the epidemic area in Uganda, together with those of EACMV-K (K2B isolate) and EACMV-T2 (Dar7 isolate). Scores represent A_{405} ranges as follows: 0 represents < 0.30 ; 1 represents $0.31-0.60$; 2 represents $0.61-1.20$; 3 represents $1.21-1.80$; and 4 represents > 1.80 . MAbs SCR 12-33 were raised against ACMV, SCR 53-66 against ICMV and SCR 105 and 106 against OLCV.

and UV-AL1/R1 and by primers UV-AL1/F1 and ACMV-CP/R3 from all 11 very severely affected cassava plants tested, which each came from a different site located up to 400 km apart in seven districts of Uganda, ranging from Lira in the north to Kampala in the south, Kasese in the south-west and Tororo in the south-east. These results show that UgV isolates were consistently found in very severely affected plants. In contrast, DNA from nine mosaic-affected plants growing outside the epidemic area of Uganda was amplified by the ACMV-specific primers but not by UV-AL1/F1 and UV-AL1/R1 or UV-AL1/F1 and ACMV-CP/R3.

The results described above provide no information on the nature of any DNA-B that might occur in plants infected with UgV isolates. However, no DNA fragment was amplified from extracts of plants containing isolates C39, C70 or C74 with two pairs of primers designed to detect ACMV DNA-B: primers ACMV-BR1/F and ACMV-BR1/R; and primers ACMV-BL1/F and ACMV-BL1/R.

Discussion

Two important results described in this paper are the discovery of a novel cassava geminivirus variant (UgV) and its association with the severe mosaic disease epidemic in Uganda. However, even more important, scientifically, is the clear evidence that the DNA-A of UgV has arisen by recombination of EACMV DNA-A and ACMV DNA-A. Recombination therefore can occur between two geminiviruses which are regarded (Hong *et al.*, 1993 and this paper), on the basis of their genomic and biological differences, as separate species. This is the first clear-cut example of such an event having occurred between identified geminivirus species. However, there are hints that recombination has played a role in the evolution of other geminiviruses. For example, most of the putative gene products of the leafhopper-transmitted beet curly top virus (BCTV), classified in subgroup (= genus) II of the

Geminiviridae, have affinities with those of whitefly-transmitted subgroup III geminiviruses whereas the CP is subgroup I-like. The BCTV genome was therefore speculated to have resulted from intergeneric recombination between a subgroup III and a subgroup I virus (Stanley *et al.*, 1986; Rybicki, 1994; Padidam *et al.*, 1995). The putative AC1 protein of pepper huasteco virus, found in Mexico, has a few amino acid residues typical of Old World and others typical of New World subgroup III geminiviruses, although the other gene products of this virus are typical of the New World viruses (Torres-Pacheco *et al.*, 1993). However, this evidence of interspecific recombination is much less strong than that provided by UgV, and the progenitor form(s) remain(s) to be identified. In contrast, evidence for interstrain recombination is convincing, as exemplified by the result of subculturing the pseudo-recombinant isolate produced by mixing DNA-A of tomato mottle virus with DNA-B of the distantly related strain known as bean dwarf mosaic virus. The sequence of the intergenic region, common to DNA-A and DNA-B, of the two parent isolates differs at several positions, but that in DNA-B of the pseudo-recombinant was converted to the form found in the DNA-A after a few serial passages in experimental plants (Hou & Gilbertson, 1996).

Our compelling evidence that geminivirus recombination can occur at the interspecific level has important implications. There are many examples of distinct species of whitefly-transmitted geminiviruses with hosts in common. Where these viruses occur in the same geographical area, numerous opportunities for recombination may exist. Recombination may be contributing, for instance, to emergence of the new forms of geminivirus which are being recorded from pepper and tomato in areas such as Central America and the southern United States (Brown & Bird, 1992). Recombination events may also have led to the general similarity in antigenic properties (Harrison *et al.*, 1991; Swanson *et al.*, 1992; Nateshan *et al.*, 1996) and genome nucleotide sequences

(Rybicki, 1994; Padidam *et al.*, 1995; Hong & Harrison, 1995) of whitefly-transmitted geminiviruses from the same geographical region but with different host ranges. Indeed, we endorse the suggestion (Rybicki, 1994) that recombination is probably a powerful factor in the evolution of whitefly-transmitted geminiviruses, not only in the long term but also in the short to medium term.

The occurrence of interspecific recombination in geminiviruses also has implications for geminivirus taxonomy and the vexed question of delineation of geminivirus species. Various pragmatic proposals have been made to assist in this objective, based either on the degree of difference of total nucleotide sequences, of sequences of encoded gene products in general or of sequences of large intergenic regions, or on the ability of two viruses to form pseudo-recombinants (Padidam *et al.*, 1995; Hong & Harrison, 1995). All these have merit. However, short cuts, such as assignment to species on the basis of the sequence of the 70 N-terminal amino acids of the CP (Padidam *et al.*, 1995), can give unreliable answers if applied to recombinant virus isolates. Thus, the special characteristics of UgV were detected only when other parts of its genome were examined (Harrison *et al.*, 1996).

Can the sequence information now available be used to decide whether UgV is best considered as a strain of the parent that provided most of its DNA-A, EACMV, or whether it should be given a distinct species name? Precedents for this situation are provided by naturally occurring tobnavirus isolates, such as N5 and I6 (Robinson *et al.*, 1987). The RNA genomes of tobnaviruses are bipartite, and in these isolates RNA-1 was typical of tobacco rattle virus (TRV) but RNA-2 contained the coat protein gene of pea early-browning virus (PEBV) inserted in an otherwise TRV-like molecule. Both N5 and I6 induced symptoms like those of TRV, and pseudo-recombinant isolates were produced by reassorting their genome parts with those of TRV but not, in most combinations, with those of PEBV. Thus, they belonged to the gene pool, and had the pathogenicity, of TRV and were considered to be TRV strains. With regard to UgV, the affinities of its DNA-B, if present, are unknown, and tests for pseudo-recombination have not been done. Moreover, its pathogenicity on a range of hosts needs to be explored. Until such information is available, we recommend that the non-committal designation, UgV, should be used.

We thank Mr J. P. Kamau and Dr J. P. Legg for providing mosaic-affected cassava material from Kenya and Tanzania. G. W. Otim-Nape is indebted to the Royal Society for the award of a Third World Fellowship. SCRI is grant-aided by the Scottish Office Agriculture, Environment and Fisheries Department.

References

- Aiton, M. M. & Harrison, B. D. (1989). Monoclonal antibodies to Indian cassava mosaic geminivirus (ICMV). *Report of the Scottish Crop Research Institute for 1988*, p. 175.
- Anon. (1994). *Program Manual for the Wisconsin Package, Version 8*. Madison: Genetics Computer Group.
- Argüello-Astorga, G., Herrera-Estrella, L. & Rivera-Bustamente, R. (1994). Experimental and theoretical definition of geminivirus origin of replication. *Plant Molecular Biology* **26**, 553–556.
- Brown, J. K. & Bird, J. (1992). Whitefly-transmitted geminiviruses and associated disorders in the Americas and the Caribbean Basin. *Plant Disease* **76**, 220–225.
- Fargette, D., Fauquet, C. & Thouvenel, J.-C. (1988). Yield losses induced by African cassava mosaic virus in relation to the mode and the date of infection. *Tropical Pest Management* **34**, 89–91.
- Fontes, E. P. B., Luckow, V. A. & Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell* **4**, 597–608.
- Gilbertson, R. L., Rojas, M. R., Russell, D. R. & Maxwell, D. P. (1991). Use of the asymmetric polymerase chain reaction and DNA sequencing to determine genetic variability of bean golden mosaic geminivirus in the Dominican Republic. *Journal of General Virology* **72**, 2843–2848.
- Harrison, B. D. & Robinson, D. J. (1988). Molecular variation in vector-borne plant viruses: epidemiological significance. *Philosophical Transactions of the Royal Society of London B* **321**, 447–462.
- Harrison, B. D., Swanson, M. M., McGrath, P. F. & Fargette, D. (1991). Patterns of antigenic variation in whitefly-transmitted geminiviruses. *Report of the Scottish Crop Research Institute for 1990*, pp. 88–90.
- Harrison, B. D., Liu, Y. L., Zhou, X., Robinson, D. J., Calvert, L. & Otim-Nape, G. W. (1996). Properties, differentiation and geographical distribution of geminiviruses that cause cassava mosaic. In *Abstracts of the Third International Meeting, Cassava Biotechnology Network (CBN III)*, 26–31 August 1996, Kampala, Uganda, p. 16.
- Harrison, B. D., Liu, Y. L., Khalid, S., Hameed, S., Otim-Nape, W. & Robinson, D. J. (1997). Detection and relationships of cotton leaf curl virus and allied whitefly-transmitted geminiviruses occurring in Pakistan. *Annals of Applied Biology* **130**, 61–75.
- Hong, Y. G. & Harrison, B. D. (1995). Nucleotide sequences from tomato leaf curl viruses from different countries: evidence for three geographically separate branches in evolution of the coat protein of whitefly-transmitted geminiviruses. *Journal of General Virology* **76**, 2043–2049.
- Hong, Y. G., Robinson, D. J. & Harrison, B. D. (1993). Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. *Journal of General Virology* **74**, 2437–2443.
- Hou, Y.-M. & Gilbertson, R. L. (1996). Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *Journal of Virology* **70**, 5430–5436.
- Macintosh, S., Robinson, D. J. & Harrison, B. D. (1992). Detection of three whitefly-transmitted geminiviruses occurring in Europe by tests with heterologous monoclonal antibodies. *Annals of Applied Biology* **121**, 297–303.
- Martin, E. F. (1928). Report of the mycologist. *Annual Report of the Department of Agriculture, Uganda*, p. 31. Entebbe, Uganda: Government Printer.
- Muniyappa, V., Swanson, M. M., Duncan, G. H. & Harrison, B. D. (1991). Particle purification, properties and epitope variability of Indian tomato leaf curl geminivirus. *Annals of Applied Biology* **118**, 595–604.
- Nateshan, H. M., Muniyappa, V., Swanson, M. M. & Harrison, B. D. (1996). Host range, vector relations and serological relationships of cotton leaf curl virus from southern India. *Annals of Applied Biology* **128**, 233–244.

- Otim-Nape, G. W., Thresh, J. M. & Fargette, D. (1996). *Bemisia tabaci* and cassava mosaic virus disease in Africa. In *Bemisia 1995: Taxonomy, Biology, Damage, Control and Management*, pp. 319–350. Edited by D. Gerling & R. T. Meyer. Andover: Intercept.
- Padidam, M., Beachy, R. N. & Fauquet, C. M. (1995). Classification and identification of geminiviruses using sequence comparisons. *Journal of General Virology* **76**, 249–263.
- Robinson, D. J., Hamilton, W. D. O., Harrison, B. D. & Baulcombe, D. C. (1987). Two anomalous tobnavirus isolates: evidence for RNA recombination in nature. *Journal of General Virology* **68**, 2551–2561.
- Rojas, M. R., Gilbertson, R. L., Russell, D. R. & Maxwell, D. P. (1993). Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Disease* **77**, 340–347.
- Rybicki, E. P. (1994). A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Archives of Virology* **139**, 49–77.
- Stanley, J. & Gay, M. R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature (London)* **301**, 260–262.
- Stanley, J., Markham, P. G., Callis, R. J. & Pinner, M. S. (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO Journal* **5**, 1761–1767.
- Swanson, M. M. (1992). *Antigenic variation and serological relationships of whitefly-transmitted geminiviruses*. PhD thesis, University of Dundee, UK.
- Swanson, M. M. & Harrison, B. D. (1993). Serological relationships and epitope profiles of isolates of okra leaf curl geminivirus from Africa and the Middle East. *Biochimie* **75**, 707–711.
- Swanson, M. M. & Harrison, B. D. (1994). Properties, relationships and distribution of cassava mosaic geminiviruses. *Tropical Science* **34**, 15–25.
- Swanson, M. M., Brown, J. K., Poulos, B. T. & Harrison, B. D. (1992). Genome affinities and epitope profiles of whitefly-transmitted geminiviruses from the Americas. *Annals of Applied Biology* **121**, 285–296.
- Thomas, J. E., Massalski, P. R. & Harrison, B. D. (1986). Production of monoclonal antibodies to African cassava mosaic virus and differences in their reactivities with other whitefly-transmitted geminiviruses. *Journal of General Virology* **67**, 2739–2748.
- Torres-Pacheco, I., Garzon-Tiznado, J. A., Herrera-Estrella, L. & Rivera-Bustamente, R. F. (1993). Complete nucleotide sequence of pepper huasteco virus: analysis and comparison with bipartite geminiviruses. *Journal of General Virology* **74**, 2225–2231.

Received 3 February 1997; Accepted 23 April 1997